

Research Article

The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy

Henning Buddenbaum¹ and Markus Steffens²

¹ *Environmental Remote Sensing and Geoinformatics, Trier University, 54286 Trier, Germany*

² *Lehrstuhl für Bodenkunde, Technische Universität München, 85350 Freising-Weihenstephan, Germany*

Correspondence should be addressed to Henning Buddenbaum, buddenbaum@uni-trier.de

Received 17 February 2012; Revised 11 May 2012; Accepted 18 September 2012

Academic Editor: Raphael Viscarra Rossel

Copyright © 2012 H. Buddenbaum and M. Steffens. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Laboratory imaging spectroscopy can be used to explore physical and chemical variations in soil profiles on a submillimetre scale. We used a hyperspectral scanner in the 400 to 1000 nm spectral range mounted in a laboratory frame to record images of two soil cores. Samples from these cores were chemically analyzed, and spectra of the sampled regions were used to train chemometric PLS regression models. With these models detailed maps of the elemental concentrations in the soil cores could be produced. Eight different spectral pretreatments were applied to the sample spectra and to the resulting images in order to explore the influence of these pre-treatments on the estimation of elemental concentrations. We found that spectral preprocessing has a minor influence on chemometry results when powerful regression algorithms like PLSR are used.

1. Introduction

Soils show a high degree of horizontal and vertical variation in physical and chemical properties. Visible and near-infrared spectroscopy is an established tool to qualitatively and quantitatively characterize these properties in soil samples [1–3]. Imaging spectroscopy is an approach that simultaneously creates VIS-NIR spectra for a complete image thus enabling analyses of the spatial distribution of these properties. In most cases imaging spectroscopy is applied from above, that is, an air- or space-borne sensor looking at the soil surface. The third spatial dimension, depth, is heterogeneous on much smaller scales but is invisible to remote sensors. Spectroscopic analyses of soil profiles can be done, for example, by measuring disturbed samples taken from different depths in the laboratory or by measuring the reflectance at different depths in boreholes [4]. However with these methods only few measurements can be made so that they cannot be used for a high-resolution characterization of complete soil profiles and their spatial variability. Our alternative is to take complete soil cores and measure their reflective properties with a laboratory imaging spectrometer [5, 6]. This way the vertical distribution of soil properties

can be studied from sub-millimetre to decimetre scale. Comparable examinations on geologic cores have been introduced by Kruse [7].

The soil core spectroscopic images can be used for various purposes, for example, for classifying soil types and their horizons [8] or for a characterization of the spatial heterogeneity of the soil profiles. This paper deals with the derivation of high-resolution maps of elemental concentrations in the soil profiles that can serve as a basis for soil classification, or for studying soil forming processes. Several regression methods (e.g., stepwise multiple linear regression [9, 10], support vector regression [10], penalized-spline signal regression [10], artificial neural networks [11, 12], multivariate adaptive regression splines [13], random forests, boosted regression trees [14], principal component regression [14, 15], narrow-band vegetation indices [16], and partial least squares regression (PLSR) [3]) have been used to quantitatively derive information from reflectance or absorption spectra. Viscarra Rossel and Behrens [17] give a comprehensive comparison of many of these techniques. Among these regression methods, PLSR [18, 19] has become one of the most popular for chemometry in recent years and will be used in this study.

In addition, spectral pretreatment can have a large impact on the result of chemometric mapping. Several spectral preprocessing methods have been introduced, for example, Ben-Dor et al. [20] used first and second derivative absorption spectra to enhance the spectral information in order to illustrate the spectral changes during a decomposition process. Udelhoven et al. [21] applied min-max normalization, convex-hull computation, first derivatives, and vector normalization, after centering each spectrum around its average. Vasques et al. [9] applied thirty preprocessing transformations, including derivatives, normalization and nonlinear transformations on spectra from 554 soil samples from Florida to derive their organic carbon content. Stevens et al. [10] used first and second derivatives, first and second gap derivatives, Savitzky-Golay smoothing and derivatives, Whittaker smoothing, standard normal variate transformation and detrending (SNV-DT), and a combination of these methods. Stenberg and Viscarra Rossel [22] show the effects of log 1/R transformation, first derivative, and SNV-DT on soil diffuse reflectance spectra. Hively et al. [23] evaluated untransformed spectra and first and second derivatives with gaps of 1 to 64 bands for the estimation of several variables from airborne hyperspectral data. Rinnan et al. [24] review the most common pre-processing techniques for near-infrared spectra in chemometry. The methods are divided into two categories: scatter-correction methods and spectral derivatives. Only few studies systematically compare the effect of these methods on chemometric spectroscopy (e.g., [9, 10, 23, 24]), and none covers pre-treatment for laboratory imaging spectroscopy. In this study, we compare the effect of eight different spectral pre-processing methods on PLSR chemometry of five chemical elements (iron = Fe, aluminium = Al, manganese = Mn, carbon = C, and nitrogen = N) in two soil profiles (Luvisol and Podzol). The pre-processing methods are as follows:

- (i) reflectance spectra (R, no pre-processing),
- (ii) standard normal variate and detrending (SNV-DT),
- (iii) first derivative of reflectance (1st D),
- (iv) second derivative of reflectance (2nd D),
- (v) continuum removed reflectance (CR),
- (vi) normalized continuum removed reflectance (NCR),
- (vii) multiplicative scatter correction (MSC),
- (viii) extended multiplicative scatter correction (EMSC).

These methods were selected because they are commonly used in spectroscopy and because each transforms the spectra in different ways with a different reasoning. For each set of transformed spectra and each chemical element PLS regression models were established; the best model was chosen and used on the image data in order to create maps of the distribution of the elements in the soil profiles. Model accuracies and maps were compared to study the effect of the spectral pre-treatments.

2. Material and Methods

2.1. Study Sites and Soil Sampling. We sampled two soil types to compare the effects of different spectral pre-treatments on the predictive power of PLSR for elemental mapping. The first profile was sampled in a Norway spruce (*Picea abies*) monoculture near Freising (SE-Germany), approximately 35 km northeast of Munich. This soil was classified as a stagnic cutanic Luvisol (ruptic, epidystric, and siltic; WRB 2006). The soil is formed on tertiary clastic sediments which is sporadically covered by quaternary aeolian sediments (loess). The second soil was sampled near Wellheim, approximately 30 km west of Ingolstadt (SE-Germany). It was classified as a folic albic Podzol (WRB 2006) that formed from cretaceous sands under a Norway spruce monoculture.

We used custom-made stainless steel boxes (100 mm × 100 mm × 300 mm) to sample 30-cm deep soil profiles. After removing the litter layer, the steel boxes were gently hammered vertically into the soil and dug out so that an undisturbed profile was sampled. The soil cores were oven-dried at 30°C to a constant weight before imaging.

2.2. Imaging Setup. The images were acquired using a hyperspectral scanner with 160 bands in the visible and near infrared 400–1000 nm spectral range (NEO HySpex VNIR-1600) mounted in a laboratory frame with a translation stage under the scanner. The translation stage moves the object in along-track direction, while the push-broom scanner records image lines across track. The speed of the translation stage is adapted so that square pixels result. The field of view in the setup is about 10 cm wide and consists of 1600 pixels resulting in about 62 μm spatial sampling distance. The full 30-cm soil cores are imaged in 4800 lines; about 200 additional image lines contain the Spectralon white reference panel and the metal frame. Light sources illuminate the currently scanned line from two directions in order to minimize shadows on the soil surface. A Spectralon white reference panel of known reflectance was scanned with the sample so radiance values could be transformed to reflectance. No geometric correction was applied to the images [5].

After recording the first image, homogeneous regions of interests (ROIs) of about 1 to 2 cm² area and about 1 cm depth were visually identified in the soil cores and sampled for chemical laboratory analysis. Samples were selected so that most of the variation within the soil cores was covered in the different samples while having a small within-sample variation. The ROIs were equally distributed over the whole profile face area and covered all horizons. After taking the samples, a layer of about 15 mm was removed from the profile face, the surface was carefully flattened, a new image of a slice parallel to the first was taken, and new ROI samples were collected. In case of the Luvisol, 7 images were recorded, and 66 samples were analyzed. Four images of the Podzol were recorded, and 35 samples were taken.

In order to explore the information content of the images, we calculated principal component analyses. Figure 4 shows an example false colour composite of principal

components of the Podzol profile, revealing information hidden in the real-colour image.

2.3. Chemical Analyses. Prior to chemical analyses, the ROI samples were dried at 50°C and sieved to <2 mm. Total C and total N concentrations were determined in duplicate by dry combustion on a EuroEA elemental analyzer (Hekatech GmbH, Wegberg, Germany). All samples were free of carbonates so that the total C concentration equals the organic carbon (OC) concentration. Quantity and quality of iron and manganese species were analysed on bulk soils <2 mm from all samples excluding the organic surface layers and organic matter rich topsoils. Total Fe and Mn oxides were extracted using the dithionite-citrate-bicarbonate-method (DCB; [25]) and measured as Fe, Mn, and Al concentrations in the extracts by inductively coupled plasma optical emission spectroscopy (ICP-OES; Vista Pro CCD Simultaneous, Varian, Darmstadt).

2.4. Spectral Pretreatments. We used 8 spectral pre-treatments prior to PLSR analyses. The mean reflectance and the transformed spectra of 35 ROI regions corresponding to the sampling spots in the Podzol profile are shown in Figure 1. Some of the techniques presented here are discussed in more details in [24].

2.4.1. Reflectance Spectra (R). Absolute reflectance ρ was derived from radiance measurements of the sample and the white reference separately for each image line by calculating the ratio of soil and reference radiances and multiplying this with the reference's known reflectance [5, 26].

In order to reduce image noise and calculation time, the image resolution was reduced by a factor of 4 (half the number of lines and rows, resp.). Then the spectra were smoothed using a Savitzky-Golay filter [27] with a 2nd-order polynomial across a moving window of 7 spectral bands. The first and last three bands were discarded, so that 154 of the original 160 bands remained. This image was used as input for the different pre-processing methods except for the derivatives. PLSR results of the spectra without further processing are the reference for the other pre-processing methods.

2.4.2. Standard Normal Variate and Detrending (SNV-DT). SNV-DT was developed by Barnes et al. [28] to remove multiplicative interferences of scatter and particle size and to account for the variation in baseline shift and curvilinearity in diffuse reflectance spectra.

Standard normal variate, also known as z-transformation or as centering and scaling [29], normalizes each spectrum ρ to zero mean and unit variance by subtracting the mean of this spectrum $\bar{\rho}$ and dividing the difference by its standard deviation σ_ρ :

$$\text{SNV} = \frac{\rho - \bar{\rho}}{\sigma_\rho}. \quad (1)$$

This is followed by a detrending step: a 2nd-order polynomial is fit to the SNV transformed spectrum and subtracted

from it to correct for wavelength-dependent scattering effects.

2.4.3. First Derivative of Reflectance (1st D). Like SNV-DT, 1st D is a method that removes the baseline from spectra while stressing absorption features. The first derivative was calculated via a Savitzky-Golay smoothing filter [11] using the hyperspectral image processing software EnMAP-Box (Version 1.1, Humboldt-Universität zu Berlin, Germany, <http://www.hu-geomatics.de/>). The original 160 band data set was used, because smoothing is part of the processing. In the Savitzky-Golay derivative procedure, a first-order polynomial was fitted to spectral windows of 7 bands width. The derivative of this polynomial was assigned as the new value of the central band. The first and last three bands were discarded, so that 154 bands resulted like in the other methods. Vasques et al. [9] consistently found Savitzky-Golay derivatives among the best pre-processing transformations. Ertlen et al. [30] state that more useful information can be extracted from near-infrared spectra if derivatives of the spectra are taken.

2.4.4. Second Derivative of Reflectance (2nd D). The 2nd D has been applied many times in remote sensing and spectroscopy, for example, for the elimination of background signals and for differentiating overlapping signatures [31, 32]. The second derivative was calculated from the 1st D spectra. Kessler [33] states that consecutive first derivatives result in less noisy spectra than higher-order derivatives directly applied to the original data, so the second derivatives were calculated as derivative of the first derivative with identical settings.

2.4.5. Continuum Removed Reflectance (CR). CR [34] is calculated by fitting a convex hull (the continuum) to the spectrum and then dividing the spectrum by the hull at each wavelength. This preprocessing gives a CR value of 1 to all parts of the spectrum that lie on the convex hull (i.e., wavelength regions that are not in an absorption band) and values between 0 and 1 to regions inside absorption bands. So CR accentuates the absorption bands in the spectra while minimizing brightness differences. Continuum removal was done in Envi (Version 4.7, ITTV is, now Exelis Visual Information Solutions). All CR calculations were applied on the complete wavelength range, not just single absorption bands.

2.4.6. Normalized Continuum Removed Reflectance (NCR). NCR spectra (also known as Band Depth Normalization, [34, 35]) were created by scaling each CR spectrum to the full 0 to 1 range by calculating

$$\text{NCR} = \frac{\text{CR} - \text{CR}_{\min}}{\text{CR}_{\max} - \text{CR}_{\min}}, \quad (2)$$

where CR_{\min} and CR_{\max} are the minimum and maximum values of a CR spectrum, respectively. The effect of band

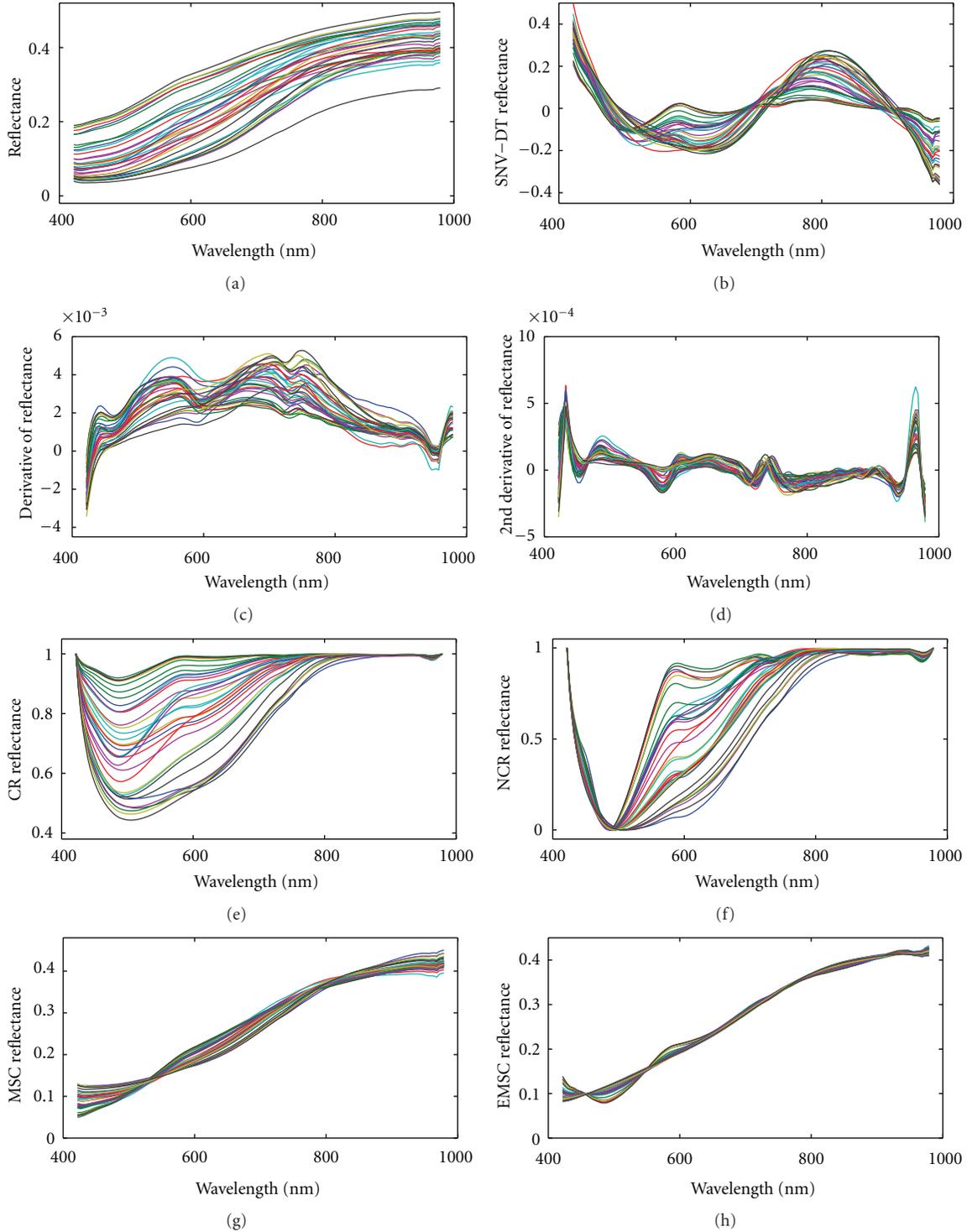


FIGURE 1: Mean reflectance (a) and preprocessed spectra ((b)–(h)) of the 35 sampled regions of interest of the Podzol soil core.

depth normalization is that the shape of absorption bands instead of their depth becomes the main feature of the spectra.

2.4.7. Multiplicative Scatter Correction (MSC). Multiplicative scatter correction [36], also known as multiplicative signal

correction, is another pre-processing technique for baseline correction in spectra. It assumes that the wavelength-dependent scatter effects on the spectrum can be separated from the chemical information. This is done by correcting the different spectra to an “ideal” spectrum so that baseline and amplification effects are at the same average level in every

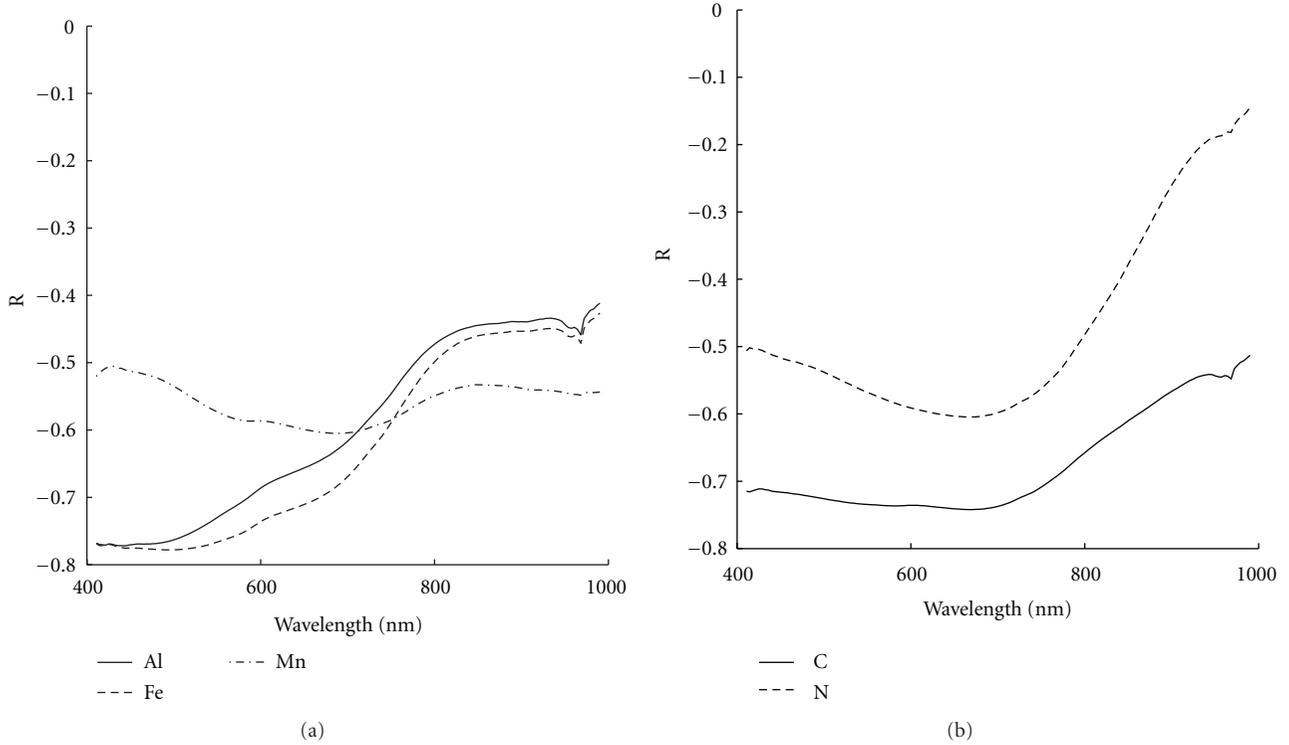


FIGURE 2: Correlation spectra for Al, Fe, and Mn content (a) and for C and N content (b) for the Podzol.

spectrum {Martens, 1991 page 156}. As this ideal spectrum is unknown, the mean spectrum \bar{x} is used. This spectrum represents the mean scattering and offset. Each spectrum x_i is then fit to the mean spectrum using a least squares method:

$$x_i = a_i + b_i \bar{x} + e_i. \quad (3)$$

Ideally, e_i contains the chemical information, because scattering and offset are represented by the coefficients a_i and b_i . The MSC spectrum is calculated by determining the coefficients for each spectrum and then transforming the spectrum as follows:

$$MSC_i = \frac{x_i - a_i}{b_i}. \quad (4)$$

2.4.8. Extended Multiplicative Scatter Correction (EMSC). MSC does not take wavelength dependences of scattering into account. EMSC extends MSC by introducing wavelength (λ) terms in order to correct for the wavelength-dependent scattering effects [33, 37]:

$$EMSC_i = \frac{x_i - a_i - d_i \lambda - e_i \lambda^2}{b_i}. \quad (5)$$

2.5. Regression Analyses with PLSR. All regression analyses were calculated using the mean reflectance or transformed spectra of the ROIs and the corresponding elemental

concentrations. The resulting regression coefficients were then applied on the reflectance or transformed images in order to create maps of the elemental concentrations. The calculations analyses were carried out in MATLAB (Version 8.0, The Mathworks).

Single reflectance bands can be correlated to elemental concentrations as measured with standard laboratory techniques. The spectral dependency of this correlation can be illustrated by a plot of the coefficient of correlation for every single band with the elemental concentration [38]. Figures 2 and 3 show correlation spectra between the reflectance values at each wavelength and the five elemental concentrations for the Podzol and the Luvisol. In the case of the Podzol, single bands have correlations of up to -0.77 , -0.78 , -0.61 , -0.74 , and -0.60 with Al, Fe, Mn, C, and N concentrations, respectively. In the case of the Luvisol, the highest single band correlations are -0.50 , -0.52 , -0.67 , -0.61 , and -0.45 . Fe oxides absorb mostly in the red spectral region, but due to the wide absorption features the correlation is high in the whole visible region, at least for the Podzol. The Luvisol has a correlation minimum around 600 nm. The Al correlation curves follow the Fe curves closely due to the high correlation between their concentrations. The Podzol Mn correlation spectrum shows no distinct features, while in the Luvisol the strongest correlation is between 450 and 500 nm. C and N have the highest correlations in the visible domain. Only in the Luvisol the C correlation is further differentiated with a correlation maximum around 600 nm. Combinations of spectral bands are known to explain higher proportions of the variance, so a tool that makes use of

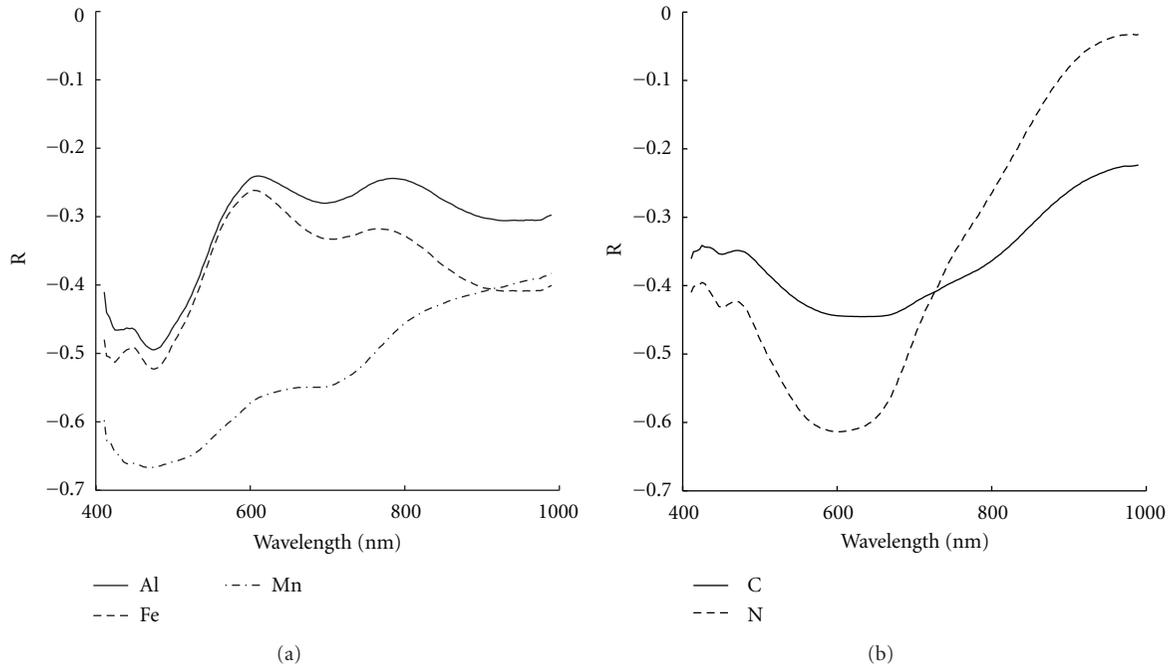


FIGURE 3: Correlation spectra for Al, Fe, and Mn content (a) and for C and N content (b) for the Luvisol.

all bands was chosen for the regression of chemical soil constituents [14].

We calculated the regression between the reflectance and the elemental concentrations with a PLSR. PLSR projects the original data into a low-dimensional space formed by a set of orthogonal latent variables by a simultaneous decomposition of X (spectral matrix) and Y (elemental concentration matrix) that maximizes the covariance between X and Y [3]. The method is well suited for the calibration of a small number of samples with experimental noise in both chemical and spectral data [14].

In order to find the optimum number of latent variables, we calculated PLSR models with 1 to 15 latent variables on the ROI spectra for each analyzed element, separately for both images. We applied leave-one-out cross-validation (LOOCV) on each model to avoid overfitting [35]. Because this was mostly a feasibility study, no further calibration/validation scheme was applied. In cases where the samples are autocorrelated, LOOCV can also increase overfitting [39], but we decided to keep the validation strategy simple because plausible maps resulted from this strategy. The accuracy of each model is given as coefficient of determination (R^2), adjusted R^2 , and relative root mean square errors (%RMSE). For each element and each spectral pre-processing method usually the number of latent variables with the lowest resulting RMSEcv was chosen. Selection of the optimal number of latent variables in the PLS estimation is a crucial step. In cases when the different measures of accuracy suggested a different number of variables, the most parsimonious model was chosen.

The adjusted R^2 is a coefficient of determination that rewards parsimonious models by incorporating the number

of regressors M and the number of observations n . It was calculated from R^2 using the following:

$$\text{adj}R^2 = 1 - (1 - R^2) \frac{n - 1}{n - M - 1}. \quad (6)$$

RMSE was calculated from the difference of predicted values y_p and observed values y_o as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{p,i} - y_{o,i})^2}. \quad (7)$$

%RMSE was derived by dividing RMSE by the mean of the observed variable.

3. Results

3.1. Chemical Analyses. Basic statistics of the chemical analyses of the two soil cores are collected in Tables 1 to 3. Some elements have a very high skewness. We repeated the calculations on log-transformed data for these elements, but the results did not get better, so we only show the results from untransformed data. Tables 2 and 3 show the correlations between the element concentrations for the Podzol and the Luvisol, respectively.

Elemental concentrations and correlations are different between the two different soil types. The Podzol has higher concentrations of Al and Fe; the Luvisol has higher concentrations of C and N. The contents of Al and Fe and those of C and N are highly correlated (Tables 2 and 3), while Mn is correlated more loosely to the other elements in both soil types. All correlations between the inorganic proxies (Al, Fe, and Mn) and the organic proxies (C and N) are positive in

TABLE 1: Basic statistics for elemental concentrations of Al, Fe, Mn, C and N in ROI samples of Podzol and Luvisol.

	Podzol					Luvisol				
	Al	Fe	Mn	C	N	Al	Fe	Mn	C	N
	[mg g ⁻¹]									
Mean	4.31	13.47	0.0139	3.05	0.076	1.736	10.844	0.873	10.33	0.599
Min	0.39	2.71	0.0043	0.46	0.010	0.997	5.796	0.241	1.797	0.180
Max	10.72	35.79	0.0252	12.53	0.260	2.451	16.397	1.885	184.7	8.505
Skew	0.457	0.628	0.238	1.67	1.41	-0.028	-0.142	0.332	2.16	6.24
Stddev	3.42	10.29	0.0059	2.46	0.066	0.382	2.664	0.470	24.23	1.088
N	35	35	33	35	16	32	32	32	66	66

TABLE 2: Correlations of elemental concentrations in the Podzol ROI soil samples with significance levels (*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$).

	Al	Fe	Mn	C	N
Al	1				
Fe	0.89***	1			
Mn	0.50**	0.62**	1		
C	0.62***	0.64***	0.38*	1	
N	0.57*	0.50*	0.28	0.93***	1

the Podzol but negative in the Luvisol. Therefore, the models for Al and Fe and the models for C and N are expected to have similar coefficients and model accuracies, respectively. These correlations also explain why Al and N can be detected by VNIR spectroscopy, although they are not optically active in the observed spectral region between 400 and 1000 nm.

3.2. Mapping of Chemical Soil Constituents. The numbers of latent variables and values of R^2 , adjusted R^2 , and %RMSE as accuracy measures achieved with these latent variables for the PLS estimations of elemental concentrations of the Podzol soil core are stated in Table 4. The corresponding results for the Luvisol soil core are shown in Table 5.

The number of latent variable for the PLS regressions is between 1 and 7 for all elements, with 4 being the most common number. Not all of the selected models explained more variance than the simple regression with a single reflectance band as explanatory variable.

Figure 4 shows a real-colour image of the Podzol profile at the left. The second panel shows principal components of the image, revealing the large amount of information in the hyperspectral image that cannot be seen by the human eye. The first principal component is not shown, because it mainly contains the brightness of the image, a piece of information that is already present in the left panel. The right panels are examples of chemometric maps of the element distribution in the profile, acquired by help of different spectral pre-treatments and application of the PLSR relations established on the ROIs to the images.

4. Discussion

The resulting submillimetre resolution maps of chemical soil constituents of 10 cm × 30 cm sections of soil profiles

(Figure 4) provide a very detailed view on the vertical soil structure. While usually spectroscopic methods are used on small, homogenized samples, or as imaging spectroscopy from above, the methods presented here facilitate new insights to the spatial distribution of elemental concentrations in soils. The maps can be used for soil classification, differentiation of characteristic horizons [8], or for the quantitative evaluation of soil forming processes.

We assume the different correlations between the organic and inorganic parameters in the two soil types (Tables 2 and 3) to be the product of different soil forming processes. In the Luvisol, C and N accumulate in the topsoil including the purely organic surface layer and the mixed organic/inorganic Ah horizon, while Al, Fe, and Mn show high concentrations in the inorganic subsoil. This spatial separation of the different materials is expressed by the negative correlation. In the Podzol, C and N accumulate together with Al, Fe, and Mn in the spodic horizon in the subsoil. This can be seen in a positive correlation between the organic and inorganic elements in the Podzol.

Figure 5 shows adjusted R^2 values for all elements and both soil types as an aggregation of Tables 4 and 5. While the amounts of Al and Fe can be estimated best in the Podzol, C and N are estimated with the highest accuracy in the Luvisol profile. C and N contents in the Luvisol are much higher than in the Podzol. This may be the reason for the higher estimation accuracies in the Luvisol. Estimations of Mn have a rather low accuracy for both soil types and all spectral pre-treatments. We assume this to be the result of the low concentrations in both soils and the associated low accuracy and the circumstance that Mn and organic substances have both low reflection across the full analysed spectral range.

The influence of the different pre-treatments on the PLSR estimation of elemental concentrations in laboratory spectroscopic images of soil profiles is rather small, especially for

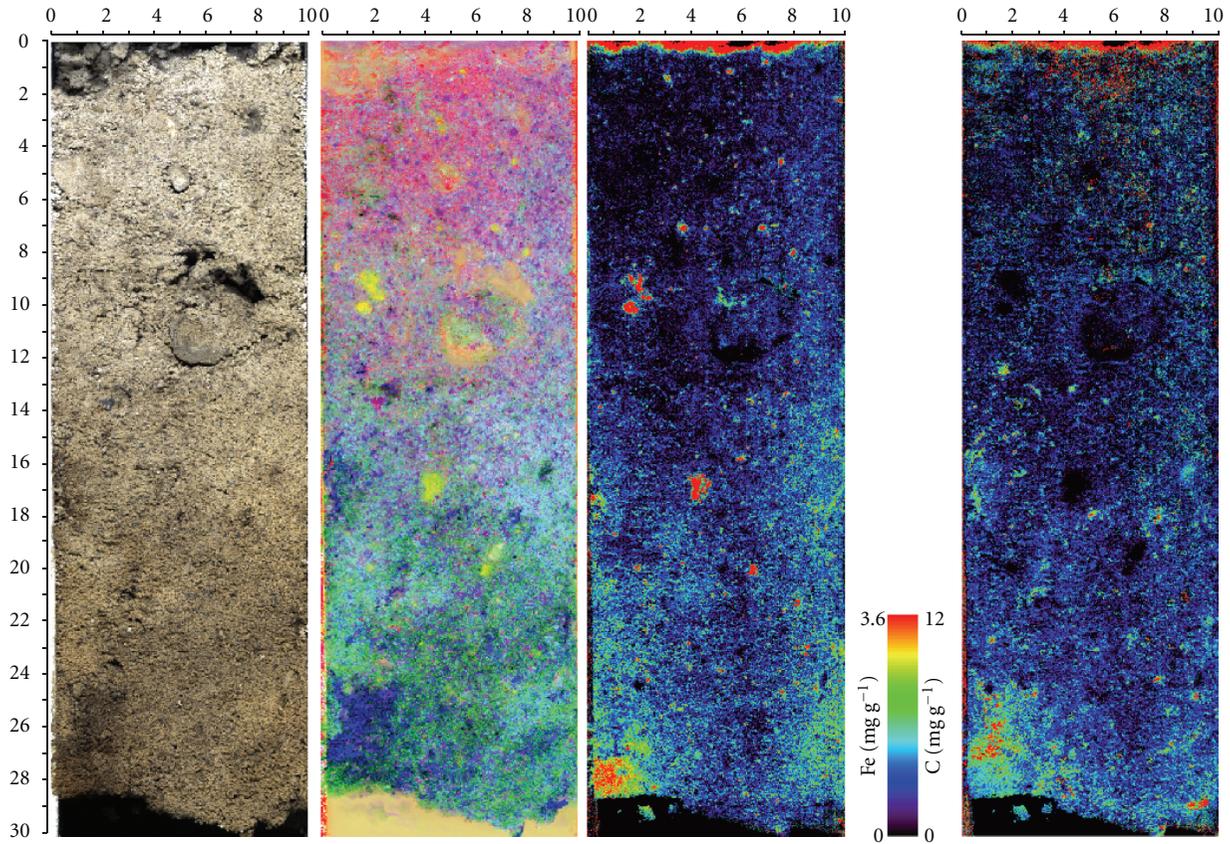


FIGURE 4: Real-colour image of the Podzol profile, false-colour image of 2nd, 3rd, and 4th principal component, map of Fe content from MSC transformed image, and map of C content from EMSC transformed image, with centimetre scales.

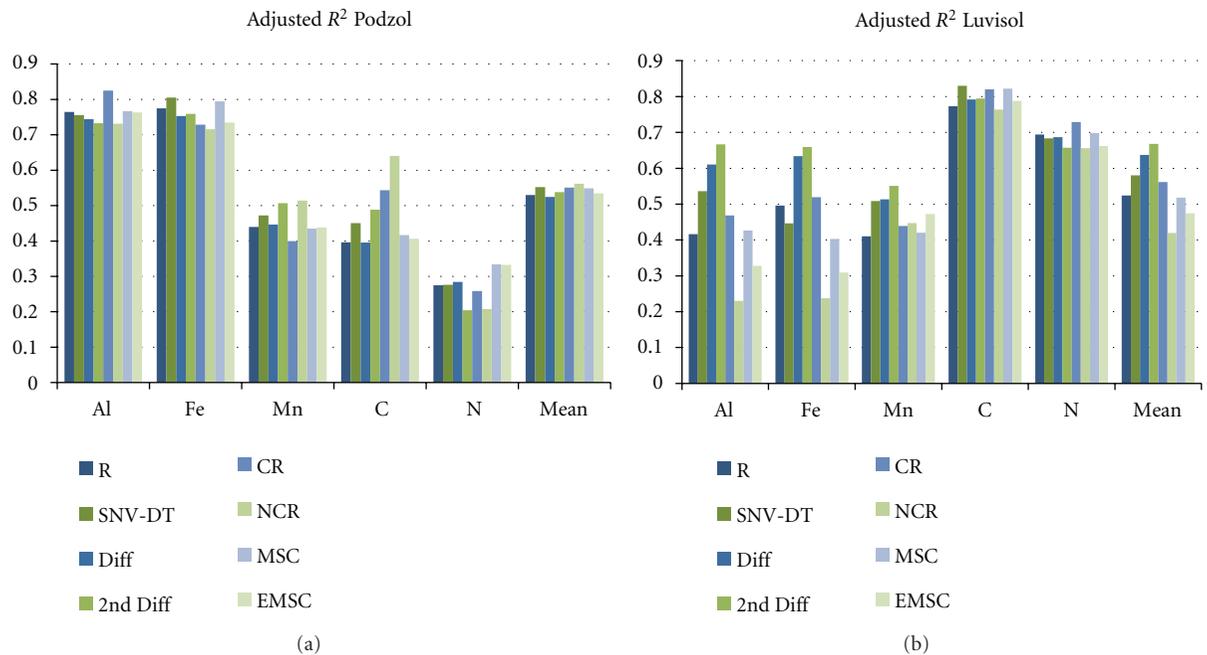


FIGURE 5: Adjusted R² values for the PLSR models of the unchanged reflectance spectra and 7 pre-treatments for Al, Fe, Mn, C, and N.

TABLE 3: Correlations of elemental concentrations in the Luvisol ROI soil samples.

	Al	Fe	Mn	C	N
Al	1				
Fe	0.96***	1			
Mn	0.65***	0.57***	1		
C	-0.39*	-0.40*	-0.51**	1	
N	-0.05	-0.04	-0.35*	0.97***	1

TABLE 4: Number of latent PLSR variables, coefficient of determination (R^2), adjusted R^2 , and relative root mean square error (%RMSE) for the selected PLSR models of chemical constituents of the Podzol.

		Al	Fe	Mn	C	N
R	N	7	7	7	5	4
	R^2	0.81	0.82	0.55	0.49	0.36
	adj. R^2	0.76	0.77	0.44	0.40	0.27
	%RMSE	34.3	32.1	28.8	58.9	75.3
SNV-DT	N	5	5	5	4	4
	R^2	0.79	0.83	0.55	0.51	0.36
	adj. R^2	0.76	0.81	0.47	0.45	0.28
	%RMSE	36.0	30.9	28.6	56.6	71.2
Diff	N	5	5	4	1	1
	R^2	0.78	0.79	0.51	0.41	0.31
	adj. R^2	0.74	0.75	0.45	0.40	0.28
	%RMSE	36.9	34.8	29.6	61.3	71.0
2nd Diff	N	4	3	4	3	1
	R^2	0.76	0.78	0.56	0.53	0.23
	adj. R^2	0.73	0.76	0.51	0.49	0.20
	%RMSE	38.8	35.3	28.0	55.3	75.2
CR	N	3	6	5	1	1
	R^2	0.84	0.78	0.49	0.56	0.28
	adj. R^2	0.82	0.73	0.40	0.54	0.26
	%RMSE	31.4	36.1	31.0	53.2	72.3
NCR	N	6	6	7	7	3
	R^2	0.78	0.77	0.61	0.71	0.28
	adj. R^2	0.73	0.72	0.51	0.64	0.21
	%RMSE	37.6	36.9	26.9	42.9	74.7
MSC	N	6	5	6	6	5
	R^2	0.81	0.82	0.53	0.52	0.43
	adj. R^2	0.77	0.79	0.44	0.42	0.33
	%RMSE	34.7	31.7	29.3	57.1	68.4
EMSC	N	5	4	8	5	4
	R^2	0.80	0.77	0.57	0.49	0.41
	adj. R^2	0.76	0.73	0.44	0.41	0.33
	%RMSE	35.4	36.8	28.1	58.1	66.2

the Podzol. Although several authors (e.g., [20, 21, 40]) note the benefit of transforming spectra before further analyses, in our case these transformations are not very helpful. Kooistra et al. [41] also found preprocessing unnecessary for the estimation of some chemicals using VNIR spectroscopy and PLSR. A reason might be that PLSR is a powerful regression technique that uses the full spectral range and thus finds the necessary information in all kinds of spectra. Spectral pre-treatment might emphasize the information contained

in the spectra, but PLSR does not seem to depend on that. Furthermore, transformations that enhance spectral features (e.g., derivatives) generally have little impact on the visible and beginning of the NIR region (i.e., the spectral region measured by the sensor), because absorption features in this region are very broad [41].

Of the pre-treatments tested, first and especially second spectral derivatives seem to be the most dangerous, although they are the most widely used methods. The derivatives

TABLE 5: Results for the Luvisol, see Table 4 for details.

		Al	Fe	Mn	C	N
R	<i>N</i>	4	4	4	5	5
	R^2	0.45	0.53	0.45	0.79	0.72
	adj. R^2	0.42	0.50	0.41	0.77	0.69
	%RMSE	16.2	16.8	39.8	41.9	32.2
SNV-DT	<i>N</i>	5	5	4	5	5
	R^2	0.57	0.49	0.54	0.84	0.71
	adj. R^2	0.54	0.45	0.51	0.83	0.68
	%RMSE	14.4	17.9	36.2	36.1	32.7
Diff	<i>N</i>	7	7	6	5	5
	R^2	0.65	0.67	0.56	0.81	0.71
	adj. R^2	0.61	0.63	0.51	0.79	0.69
	%RMSE	13.1	14.2	36.4	40.0	32.6
2nd Diff	<i>N</i>	5	6	5	3	2
	R^2	0.69	0.69	0.59	0.80	0.67
	adj. R^2	0.67	0.66	0.55	0.79	0.66
	%RMSE	12.2	13.6	34.8	40.4	34.9
CR	<i>N</i>	3	3	4	2	2
	R^2	0.49	0.54	0.47	0.83	0.74
	adj. R^2	0.47	0.52	0.44	0.82	0.73
	%RMSE	15.6	16.6	38.9	38.1	31.0
NCR	<i>N</i>	2	2	3	2	2
	R^2	0.25	0.26	0.47	0.77	0.67
	adj. R^2	0.23	0.24	0.45	0.76	0.66
	%RMSE	18.9	21.1	38.9	43.7	35.0
MSC	<i>N</i>	3	3	3	3	4
	R^2	0.45	0.43	0.45	0.83	0.72
	adj. R^2	0.43	0.40	0.42	0.82	0.70
	%RMSE	16.2	18.5	39.7	37.6	32.2
EMSC	<i>N</i>	3	3	3	3	3
	R^2	0.36	0.34	0.50	0.80	0.68
	adj. R^2	0.33	0.31	0.47	0.79	0.66
	%RMSE	17.6	20.0	37.8	41.0	34.3

emphasize noise in the data more distinctly than the other methods, so they should only be used when a very low noise level is ensured, either by low noise data or by filtering the data before or during the calculation of the derivatives. But still, they lead to the best estimations of elemental concentrations in the Luvisol. NCR is the least recommendable method for our application of quantitatively deriving elemental concentrations. Usually the concentration is linked to band depths, so normalizing band depths eliminate parts of the desired information. This is reflected in generally low accuracy values from NCR spectra. The CR estimations have the same accuracy as the estimations from untreated R spectra. CR is commonly used for baseline corrections, that is, especially differences in illumination and in viewing geometry. In our case, smooth surfaces and artificial light from two directions combined with a column-wise radiometric correction of the images resulted in very uniform illumination. This might explain the small benefit of CR. The same is probably true for SNV-DT, MSC, and EMSC. These

pre-treatment methods are designed for baseline corrections, so their benefit is small if the baseline does not vary much.

Limitations of this study are that only two different soil profiles were analyzed, that only a limited number of samples were available, and that only the wavelength region of 400 to 1000 nm was considered. The relatively high variability of chemical soil constituents in the limited space of the soil cores considered made it possible to train regression models with acceptable accuracies that could be used for creating maps of the vertical distribution of chemical soil constituents in a very high spatial resolution. Future work should include several soil profiles from the same area to be able to make robust claims on the vertical distribution of soil properties in that area.

5. Conclusions

Laboratory imaging spectroscopy was used for mapping the small-scale distribution of elemental concentrations

in soil profiles. PLSR is a powerful regression tool that makes use of all input bands and served well in finding the optimal combination of spectral bands representing specific elemental concentrations. Eight different spectral pre-treatments were tested but not deemed necessary for PLSR analyses and only in some of the cases increased the prediction accuracy of the PLSR. The estimation accuracy of the different elemental concentrations varies according to their optical activity and their concentration. Furthermore, there are no global predictors for elemental concentrations across different soil types, and the analyses have to be adjusted to the given conditions. In future studies we plan to extend the spectral range of soil profile imaging spectroscopy to the short-wave infrared region of 1000 to 2500 nm. Since many absorption bands lie in this spectral region, even better chemometric mapping is expected from this.

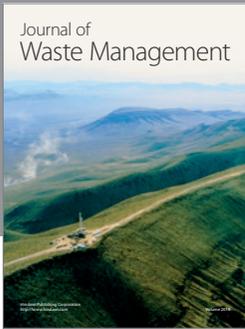
Acknowledgments

Hans and Florian Steffens are gratefully acknowledged for the technical assistance and Joachim Hill from the Department of Environmental Remote Sensing and Geoinformatics at the University of Trier for providing the imaging spectrometer. The authors are grateful to three anonymous reviewers and the editor who gave valuable comments and suggestions. This research was supported within the framework of the EnMAP project (Contract no. 50EE0946-50) by the German Aerospace Center (DLR) and the Federal Ministry of Economics and Technology.

References

- [1] E. R. Stoner and M. F. Baumgardner, "Characteristic variations in reflectance of surface soils," *Soil Science Society of America Journal*, vol. 45, no. 6, pp. 1161–1165, 1981.
- [2] E. Ben-Dor, S. Chabrilat, J. A. M. Dematté et al., "Using Imaging Spectroscopy to study soil properties," *Remote Sensing of Environment*, vol. 113, pp. S38–S55, 2009.
- [3] M. Vohland and C. Emmerling, "Determination of total soil organic C and hot water-extractable C from VIS-NIR soil reflectance with partial least squares regression and spectral feature selection techniques," *European Journal of Soil Science*, vol. 62, no. 4, pp. 598–606, 2011.
- [4] E. Ben-Dor, D. Heller, and A. Chudnovsky, "A novel method of classifying soil profiles in the field using optical means," *Soil Science Society of America Journal*, vol. 72, no. 4, pp. 1113–1123, 2008.
- [5] H. Buddenbaum and M. Steffens, "Laboratory imaging spectroscopy of soil profiles," *Journal of Spectral Imaging*, vol. 2, pp. 1–5, 2011.
- [6] H. Buddenbaum and M. Steffens, "Mapping the distribution of chemical properties in soil profiles using laboratory imaging spectroscopy, SVM and PLS regression," *EARSeL EProceedings*, vol. 11, no. 1, pp. 25–32, 2012.
- [7] F. A. Kruse, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *International Journal of Remote Sensing*, vol. 17, no. 9, pp. 1623–1632, 1996.
- [8] R. A. Viscarra Rossel and R. Webster, "Discrimination of Australian soil horizons and classes from their visible-near infrared spectra," *European Journal of Soil Science*, vol. 62, no. 4, pp. 637–647, 2011.
- [9] G. M. Vasques, S. Grunwald, and J. O. Sickman, "Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra," *Geoderma*, vol. 146, no. 1-2, pp. 14–25, 2008.
- [10] A. Stevens, T. Udelhoven, A. Denis et al., "Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy," *Geoderma*, vol. 158, no. 1-2, pp. 32–45, 2010.
- [11] A. M. Mouazen, B. Kuang, J. De Baerdemaeker, and H. Ramon, "Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy," *Geoderma*, vol. 158, no. 1-2, pp. 23–31, 2010.
- [12] J. Farifteh, F. Van der Meer, C. Atzberger, and E. J. M. Carranza, "Quantitative analysis of salt-affected soil reflectance spectra: a comparison of two adaptive methods (PLSR and ANN)," *Remote Sensing of Environment*, vol. 110, no. 1, pp. 59–78, 2007.
- [13] K. D. Shepherd and M. G. Walsh, "Development of reflectance spectral libraries for characterization of soil properties," *Soil Science Society of America Journal*, vol. 66, no. 3, pp. 988–998, 2002.
- [14] C. Atzberger, M. Guérif, F. Baret, and W. Werner, "Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat," *Computers and Electronics in Agriculture*, vol. 73, no. 2, pp. 165–173, 2010.
- [15] C. W. Chang, D. A. Laird, M. J. Mausbach, and C. R. Hurburgh, "Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties," *Soil Science Society of America Journal*, vol. 65, no. 2, pp. 480–490, 2001.
- [16] M. Schlerf, C. Atzberger, and J. Hill, "Remote sensing of forest biophysical variables using HyMap imaging spectrometer data," *Remote Sensing of Environment*, vol. 95, no. 2, pp. 177–194, 2005.
- [17] R. A. Viscarra Rossel and T. Behrens, "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma*, vol. 158, no. 1-2, pp. 46–54, 2010.
- [18] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [19] S. Wold, J. Trygg, A. Berglund, and H. Antti, "Some recent developments in PLS modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 131–150, 2001.
- [20] E. Ben-Dor, Y. Inbar, and Y. Chen, "The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process," *Remote Sensing of Environment*, vol. 61, no. 1, pp. 1–15, 1997.
- [21] T. Udelhoven, C. Emmerling, and T. Jarmer, "Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study," *Plant and Soil*, vol. 251, no. 2, pp. 319–329, 2003.
- [22] B. Stenberg and R. A. Viscarra Rossel, "Diffuse reflectance spectroscopy for high-resolution soil sensing," in *Proximal Soil Sensing*, R. A. Viscarra Rossel, B. A. McBratney, and B. Minasny, Eds., pp. 29–47, Springer Science+Business, Dordrecht, The Netherlands, 2010.
- [23] W. D. Hively, G. W. McCarty, J. B. Reeves et al., "Use of airborne hyperspectral imagery to map soil properties in tilled

- agricultural fields,” *Applied and Environmental Soil Science*, vol. 2011, Article ID 358193, 13 pages, 2011.
- [24] Å. Rinnan, F. V. D. Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” *Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.
- [25] O. P. Mehra and M. L. Jackson, “Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate,” in *Proceedings of the 7th National Conference on Clays and Clay Minerals*, pp. 317–327, 1960.
- [26] D. R. Peddle, H. P. White, R. J. Soffer, J. R. Miller, and E. F. LeDrew, “Reflectance processing of remote sensing spectroradiometer data,” *Computers and Geosciences*, vol. 27, no. 2, pp. 203–213, 2001.
- [27] A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least squares procedures,” *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [28] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, “Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra,” *Applied Spectroscopy*, vol. 43, no. 5, pp. 772–777, 1989.
- [29] O. Otto, *Statistics and Computer Application in Analytical Chemistry*, Wiley-VCH, Weinheim, Germany, 1998.
- [30] D. Ertlen, D. Schwartz, M. Trautmann, R. Webster, and D. Brunet, “Discriminating between organic matter in soil from grass and forest by near-infrared spectroscopy,” *European Journal of Soil Science*, vol. 61, no. 2, pp. 207–216, 2010.
- [31] B. L. Becker, D. P. Lusch, and J. Qi, “Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis,” *Remote Sensing of Environment*, vol. 97, no. 2, pp. 238–248, 2005.
- [32] Y. Li, T. H. Demetriades-Shah, E. T. Kanemasu, J. K. Shultis, and M. B. Kirkham, “Use of second derivatives of canopy reflectance for monitoring prairie vegetation over different soil backgrounds,” *Remote Sensing of Environment*, vol. 44, no. 1, pp. 81–87, 1993.
- [33] W. Kessler, *Multivariate Datenanalyse Für Die Pharma-, Bio. Und Prozessanalytik*, Wiley-VCH, Weinheim, Germany, 2007.
- [34] R. F. Kokaly and R. N. Clark, “Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression,” *Remote Sensing of Environment*, vol. 67, no. 3, pp. 267–287, 1999.
- [35] M. Schlerf, C. Atzberger, J. Hill, H. Buddenbaum, W. Werner, and G. Schüller, “Retrieval of chlorophyll and nitrogen in Norway spruce (*Picea abies* L. Karst.) using imaging spectroscopy,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 12, no. 1, pp. 17–26, 2010.
- [36] B. Datt, “Visible/near infrared reflectance and chlorophyll content in Eucalyptus leaves,” *International Journal of Remote Sensing*, vol. 20, no. 14, pp. 2741–2759, 1999.
- [37] H. Martens and E. Stark, “Extended multiplicative signal correction and spectral interference subtraction: new pre-processing methods for near infrared spectroscopy,” *Journal of Pharmaceutical and Biomedical Analysis*, vol. 9, no. 8, pp. 625–635, 1991.
- [38] M. Vohland, C. Bossung, and H. C. Fründ, “A spectroscopic approach to assess trace—heavy metal contents in contaminated floodplain soils via spectrally active soil components,” *Journal of Plant Nutrition and Soil Science*, vol. 172, no. 2, pp. 201–209, 2009.
- [39] D. J. Brus, B. Kempen, and G. B. M. Heuvelink, “Sampling for validation of digital soil maps,” *European Journal of Soil Science*, vol. 62, no. 3, pp. 394–407, 2011.
- [40] R. A. V. Rossel and C. Chen, “Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils,” *Remote Sensing of Environment*, vol. 115, no. 6, pp. 1443–1455, 2011.
- [41] L. Kooistra, R. Wehrens, R. S. E. W. Leuven, and L. M. C. Buydens, “Possibilities of visible-near-infrared spectroscopy for the assessment of soil contamination in river floodplains,” *Analytica Chimica Acta*, vol. 446, no. 1-2, pp. 97–105, 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

